智力科学研究及其入门问题

作者：顾立帆

一、引言

近年人工智能在围棋、无人汽车、自然语言处理、分子结构和制药等等研究领域有些进展。但缺乏准确、严谨的分析、测试方法，使这些进展被明显夸大，可能误导未来科学发展。

计算机围棋、无人汽车、自然语言处理跟人类智力密切交互。需要对高度不稳定的人类智力作结构性、系统性研究，才能有效设计科学实验、准确检测评估这些技术、找出潜在问题。

语言涉及判断力，是数学、哲学、科学的基础，也是文明的基础。由于各种计算技术（包括量子计算）均无法准确处理高阶逻辑、抽象思维、语义演变发展和诡辩分析，可以设计实验找出语言处理中人工智能的上限。而设计实验在围棋中找出人工智能的更低上限则更具挑战性。

由于哥德尔不完备定理、海森堡测不准原理和熵定律趋势，人工智能和各种计算技术有严重局限性。Universal Approximation Theorem 和 Nash Embedding Theorems 的条件设定对很多领域不适用。智力上的高度不稳定性和其它问题，需要新逻辑体系和新实验规范提供研究线索和严谨基础。

批判（critique）一词，起源于拉丁文 criticus 和希腊文 κριτικός，跟 critical 和 judgement 有关。学术批判也是发现对称性破缺、寻找逻辑临界点并形成更高层次抽象、更高境界理性的关键。

而自由意志是否存在、如果存在跟量子随机性、基因、自我意识有什么本质区别，引入何种自由意志会对人类直觉思维和学术批判能力产生积极影响，会给智力模型的逻辑、结构、功能带来什么进化，会对各种人工智能技术取得什么优势？这些是智力科学研究的关键。

根据这些问题的研究，笔者分析了各种人工智能技术的局限和问题；提出能综合自由意志、人类直觉思维、应对熵定律的各种方法、诡辩分析的新逻辑体系；以此为基础设计测试计算机围棋的四维实验空间和测试语言处理的数据集，并整理二十八道智力研究入门问题用于未来教育。

二、人工智能技术的局限和问题

基于一阶逻辑的各种数学和计算技术缺乏足够抽象程度描述人类特有智力。深度学习对高阶逻辑的模仿在模式演变下很不稳定。数字化方法会破坏很多重要结构和系统信息。跟人类智力密切交互的自动系统可以被人为诱发、在模式不稳定演变下会自动放大这些问题。

现有人工智能测试，包括 Turing Test、AlphaGo 系列的对弈、人工智能参加为人类设计的考试、语言处理的各种 dataset、对无人汽车的各种模拟测试和路试等等，都是经验测试。不但缺乏理论分析以控制实验条件来突破搜索局限、有效分离各种因素、准确找出不稳定演变的触发条件和演变原理，也无法正确分析和解释实验结果，远远达不到科学实验和科学结论的严谨标准。

如计算机围棋和人类的对局及计算机围棋之间的对局，占围棋所有可能空间均微不足道，也缺乏经过严格验证的可靠原理来推导尚未探索过的情况。

笔者可以证明：即使在围棋上，人工智能也有上限。笔者设计了四维围棋实验空间来扩大搜索范围，证伪 AlphaGo Zero 的 superhuman 宣称，验证 AlphaGo 系列（包括 AlphaZero、MuZero 和其它可能后继）缺乏人类直觉思维能力、智力现象在模式演变上的高度不稳定性、人类顶级棋手能否对最强计算机围棋在实验空间取得优势或拉锯，并找出人工智能较精确的上限和针对具体人工智能结构的更精确上限，等等。

根据对自由意志的分析，笔者就可以证明无人汽车 SAE level 4 不可能实现。当然还有更严重的问题。无人汽车的模式不稳定演变也更加复杂，其模拟测试和路试占所以可能空间更微不足道。很多严重问题现在不会出现但以后会随模式的不稳定演变大概率出现。

很多年前就听说无人汽车技术已经 ready，只是法律落后。带头试图实用化的 ALphabet/Waymo 无人汽车现在能否达到 SAE level 4？这个问题应该由谁决定、怎样有效检测？对这样关系到公众安全的重要问题，如不能准确判定，是否 SAE level 4 标准制定得不完善？如对 ALphabet/Waymo 无人汽车无法准确测试、评估，其他厂商会以此为借口进一步放松安全底线。

问题原因在于没有找到有效观察量。智力科学需要引入有效观察量概念，并根据新逻辑体系对有效观察量概念作重新定义。根据有效观察量，就可以准确地制定测试方法和评估标准。

Reward Is Enough 观点的错误在于：reward 不是有效观察量，各种短期 reward 和长期 reward 可能指向不同、甚至相反方向。针对不同场景情况更加复杂。Reward 还可能无法定义、测量，甚至无法被观察，因此根本无法被人工智能技术利用。

计算机语言处理有很复杂问题。虽然人工智能技术已能处理很多常见语言问题，但很多不常见语言问题更重要，却无法由人工智能处理。这些涉及学术研究、教育、医疗、司法等等，甚至关系到科学和文明兴衰。很多这样的问题涉及高阶逻辑、抽象思维、语义演变发展和诡辩分析。像上述 SAE level 4 这样的工业标准问题，也无法由人工智能解决，只能由人类制定修改处理。

事实上，Nash Embedding Theorems 有局限性，无法作为逻辑临界点分析的数学基础。所以，计算机语言处理中的 Word-embedding Vector Space 也无法处理对科学和文明兴衰关系重大的逻辑和诡辩问题。语言处理和人工智能研究需要新的数学理论和方法。

由于缺乏对语言智力的系统、结构研究，现有测试计算机语言处理的数据集无法测试关键的逻辑和诡辩分析以及其它很多重要问题，大大高估了人工智能的智力水准。

目前药物研发不得不较多使用猜测。在人类研究者所作前期研究已积累大量有效数据的特殊情况下，利用人类知识的人工智能可能比人类猜得更准。但在药物成败的关键：漫长的临床试验上，人工智能未必有多少帮助。一旦人类在药物原理上取得重要进展，人工智能在中期的猜测作用也会消失。而不依赖人类知识的人工智能对分子结构和药物的分析研究有多少帮助更是问题。

人工智能技术的很多问题，涉及计算技术（包括量子计算）的本质缺陷。而神经科学仅在生理层次的动物性智力：视觉、声觉、运动、情绪、记忆、经验、自我意识等等方面有些进展。无法真正研究人类特有智力如语言、数学、哲学、科学的发展。以神经科学和生命科学为基础的深度学习（deep learning）和强化学习（reinforcement learning）也有这样的局限。

科学发展早已超出直观的范围，进一步发展新科学的线索往往是特殊情况下出现的意外和罕见数据。深度学习不但缺乏对逻辑的敏感性和准确性来分析这种新线索的含义，还会淹没这种罕见的新线索。强化学习根本无法理解新线索和设计证伪实验的价值及对未来科学发展的 reward。所以深度学习和强化学习不但不是科学方法，还会误导科学发展。排斥证伪实验也是对科学的误导。

科学不同于数学；生命科学不同于物理科学；智力科学也不同于生命科学，关注的有效观察量不同，其哲学基础和思维方法也不同。下一节介绍笔者的智力科学理论研究和实验方法。

三、笔者的智力理论研究和实验方法

生命科学只能研究演化，无法研究人类特有智力的进化和退化。由于熵定律的影响，生命实际上在逆水行舟，很容易出现智力退化。智力科学需要不同于生命科学的观察量和实验方法。

很多人陷在牛顿世界观和拉普拉斯妖的错误观念中，Universal Approximation Theorem 仍延续这种错误，无法模拟人类特有智力。需要研究自由意志是否存在，如果存在跟量子随机性、基因、自我意识有何本质区别，对人类直觉思维和学术批判能力有什么作用？即什么思维能力能摆脱量子随机性、selfish gene 的表达、感官接受的数据信息所形成的自我意识等等的局限？

语言是人类重要的智力特征。其他已知生物，都不具有复杂的语言能力。在语言基础上，人类发展出两类重要的学术：数学和哲学。数学向精确性发展，哲学向完整性发展。

[illegible]

[illegible]

[illegible] logical positivism [illegible] logical empiricism [illegible] Positivism [illegible] empiricism [illegible]

[illegible] Category Theory [illegible] critique [illegible]

[illegible] critique [illegible] critique [illegible]

[illegible]

[illegible] IT [illegible]

[illegible]

1 [illegible] AlphaGo [illegible]

Deepmind [illegible] AlphaGo Zero [illegible] AlphaGo [illegible]

2 [illegible]

3 [illegible]

4 [illegible] Axiom of Choice [illegible]

[illegible] 1) [illegible] 2) [illegible] 3) [illegible] 4) [illegible] 1) [illegible] 2) [illegible]

[illegible]

Leukotomy 如此严重的学术错误七十多年无法纠正，说明 selfish gene 能实现公益的理论错误。而 Technological Singularity、potentially a meta-solution to any problem、Reward Is Enough 等等问题观点流行，说明智力研究和人工智能测试、liberal arts 和前沿研究性大学的通识教育都需要变革。

罗马俱乐部的人口资源理论和现在的气候变暖理论，都是数学推论，无法用实验验证，只能靠时间检验。而笔者的智力理论，可以用实验验证。这是科学和数学的区别。笔者不反对气候变暖理论，只是说它达不到科学标准。而人口、经济发展不仅仅是数学，跟人文、智力密切相关。

四、智力研究的入门问题

笔者整理的二十八个智力研究入门问题大致上可分成 A、B、C、D 四部分。问题如下：

A. 数和语言的基本结构问题：

1. 科学是否需要无理数，或对此尚不清楚、需进一步研究？如果回答不需要无理数，请给出充分理由。

2. 请完整表述无理数的语义。

3. 请找出 Chaitin's constant 语义上的问题及其对智力发展的特殊影响和争议。

4. 请构造一个例子说明不可定义数概念不稳定。

5. 请根据问题 1 - 4 抽象出语言处理的基本问题，并说明为何这是基本问题。

B. 人类直觉和统计及关联性的不同：

6. 请构造例子说明统计方法会失效并引起严重误导；并分析关联理论（relevance theory ）所带来的价值以及在抽象、分析、判断等等方面的不足。

7. 有些群体在历史上尚处于原始部落状态时就已发展出非常复杂的构词法规则，但在当时并没有形成更灵活、表达力更丰富而又严谨的句法。请分析语言变革和发展的需要，以及在其背后的智力结构和演变规律。

8. Grigori Perelman 的 Poincaré conjecture 论文并非用形式逻辑写成。请问除了形式逻辑，数学研究和证明在本质上还需要什么重要思维方法？请尽可能用数学语言完整描述这些思维方法，并说明语言、数学、哲学、科学研究中所用直觉各自需要什么严谨规范。

9. Demis Hassabis 把 AlphaGo 系列中的智能称为 intuition。请问数学研究中所用到的 intuition 方法跟 Demis Hassabis 所说的 AlphaGo 系列中的 intuition 有什么区别，类似 AlphaGo 系列的技术有没有可能成为 a meta-solution to any problem？

10. 请问 AlphaGo 系列是否如其 Nature 论文所说已达到 superhuman performance，是否是围棋的最优解？请证明或设计实验验证你的答案，并说明验证普适性或广适性判断需要满足什么条件。

C. 广义的语言、人性、经济和自由意志：

11. 中国古代诗词经历了格律、词牌、曲牌的形成、繁荣、然后被打破，欧洲近代音乐也经历了各种 form 的形成、繁荣、然后被打破。请问：这种范式的形成、繁荣、然后被打破，背后有什么智力规律在驱动、起了什么作用、其适用有什么特殊限制？

12. 贝多芬第五交响乐一开始的 motif 被广泛解释成命运的叩门声。但也有说法说是黄鹂叫声。请问能不能否定黄鹂叫声的说法。如果能，请说明充分

的理由。如果不能，请构造黄鹂叫声说法的合理解释。

13. 瓦格纳在音乐中想要表达的“truth”是不是真的 truth，怎么来验证？为什么他的作品中一再出现用药的情节以及用药来解释情节发展？人性和人格发展的真正逻辑是什么？可能会导致诸神的黄昏的真正问题又是什么？

14. 请从智力原理上分析 The Selfish Gene（或 The Immortal Gene）理论的问题在哪里？这跟瓦格纳的音乐有什么关系？

15. 艺术作品常用隐喻来描述人格特征，如白天鹅、黑天鹅。Freeman Dyson 的文章 Birds and Frogs 也用 birds 和 frogs 来描述研究者的思维特征。这种用借代的具体形象来暗示抽象特征的语义难以用计算机图像处理或语言处理来获得。请分析类似的各种语言手法（包括音乐和视觉艺术手法）背后的智力特征及其规律。

16. 请分析奥地利经济学派（Austrian School of Economics）对社会的影响和其尚未解释清楚的问题，以及各种现代经济学派的不足。

17. 上帝是否掷骰子已有人回答。请问人类是否有自由意志？如果没有，您是否只是由物理规律操纵的傀儡？如果自由意志存在，跟量子随机性、selfish gene 的表达、自我意识有什么区别，对发展人类特有智力、特别是探索新科学有什么作用？是否可以测量、是否可以用物理规律表达、对物理科学中的各相对性原理有没有影响、对与人类智力密切交互的人工智能、无人汽车等等技术又有什么影响？

D. 对智力结构和系统的进一步分析和综合**:**

18. 请尽可能完整地定义人类特有智力，并说明你的定义能给智力科学研究带来什么重要价值。

19. 亚里斯多德建立逻辑学试图清除诡辩。但哥德尔的研究发现逻辑本身也有问题。请问人类应该怎样处理诡辩及其危害，计算机包括量子计算机能否做到这一点？

20. 请分析为何“白马非马”在通常语境中是诡辩，而“道可道，非常道；名可名，非常名”却是重要哲学原理。请进一步分析观察方法的局限性和逻辑实证主义的问题，并基于科学只能证伪的哲学原理提出解决方法。

21. 请分析基于一阶逻辑的数学在表达人类特有智力的抽象、分析、判断等等能力上的不足和缺陷。并分析这些问题给电子计算机、量子计算机以及各种 Turing Machine 模型（deterministic, probabilistic, etc.）所带来的本质局限性。

22. 请问 Turing Test 是否有效。如果无效，请说明充分理由。并分析对自动驾驶 SAE level 4 和 level 5 的验证需要什么必要条件来排除未来不稳定模式演变所带来的大概率严重问题。

23. 请分析各种计算机语言模型和处理方法的本质缺陷（包括基于 word-embedding vector space 和 encoder-decoder, attention, transformer, BERT 等等技术的）、现有语言处理测试方法的不足和偏差、并提出测试方法的改进。请分析数学、哲学、科学的发展对语言发展有什么要求？

24. 请分析并找出各种 deep-learning 架构（包括 deep residual networks 及其变种，generative adversarial networks, etc.）所缺失的人类特有智力因素以及所造成的人工智能缺陷。

25. 请分析跟科学方法相比，用 Universal Approximation Theorem 概括的各种数据处理方法有什么缺陷、能否模拟人类特有智力；并分析按科学标准 overfitting，underfitting 等等术语能否描述人类特有智力。请找出 chaos phenomena 中需要解决的相关问题。

26. 请找出 reward 无法定义、可以定义但无法测量、可以测量但无法被观察的相应例子，并分析 Reward Is Enough 观点的问题。

27. [illegible]selfish gene [illegible]

28. [illegible]

[illegible]

[illegible] Freeman Dyson [illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible] AlphaGo [illegible] Nature [illegible] SAE level 5 [illegible]SAE level 4 [illegible]

[illegible]

[illegible]

[illegible]

[illegible]The Selfish Gene [illegible]

[illegible] Freeman Dyson [illegible] a great bird[illegible] frog[illegible] bird [illegible] frog [illegible] frog [illegible] frog [illegible] bird[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible] natural law [illegible]

[illegible]

[illegible]

[illegible]

Deepmind [illegible] Reward Is Enough[illegible] Reward Is Enough [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

reward